# PART 2: GUIDELINES FOR PRESENTING DATA AND STATISTICS IN TABLES, CHARTS, AND GRAPHS

"Data should be tabulated only when it is possible to present them more concretely in that form than in the text of the article. A cumbrous and unwieldly table defeats the purpose for which it was intended."[1]

"A picture *is* worth a 10,000 words.
But it uses up 10,000 times the memory."[2]

Tables, charts, and graphs are often used to report data and statistics. Each of these 3 graphic devices conveys information by showing 1) values and 2) labels in a 3) context. *Values* are data; they may be numbers, symbols, or text. Values by themselves are meaningless: 120, +, or "mild," for example. Thus, they must be accompanied by labels. *Labels* identify the values and often add additional characteristics: in the above examples, a mean systolic pressure of 120 mm Hg; a positive diagnostic test result; and "mild" symptoms of Alzheimer's disease. Even values with labels, however, are meaningless in the absence of a larger context, such as a specific research project. *Context* is the background against which values and labels should be interpreted. To continue the above example, a mean systolic pressure of 120 mm Hg might be the therapeutic endpoint for patients with hypertension; a positive diagnostic test result might define a subgroup of patients with comorbidities; and "mild" symptoms of Alzheimer's disease might be an inclusion criterion in a drug study. In tables, charts, and graphs, the context should be identified in the titles or captions, as well as be apparent from the larger document.

These 3 graphic devices are also similar in that all involve a coordinate system to convey additional information about values. **Tables** are good for presenting categorical data measured with nominal, binomial, or ordinal values, or with *summary statistics* of continuous values (means, medians, ranges, and so on). Here, the column and row heads label the values in each cell. **Charts** are good for presenting categorical data with values measured on either an ordinal or a continuous scale or with summary statistics of continuous values. In bar, column, or dot charts, the categories are listed (labeled) on one axis and the value for the category is given by the other axis. Charts are thus an intermediate step between tables and graphs. **Graphs** are good for presenting data with values usually measured on a continuous scale. Here, the values are read from the 2 (or sometimes 3) scales that, in addition, label the values.

In summary, and among other characteristics described in the next 2 chapters, good tables, charts, and graphs should 1) present values clearly, 2) label these values appropriately, and 3) provide enough context for readers to interpret the values correctly.

---

1. Simmons GH, Fishbein M. The Art and Practice of Medical Writing. Chicago: American Medical Association, 1925.

2. Unknown cyberhumorist

# CHAPTER 17
## TABULAR DISPLAYS OF DATA AND STATISTICS: REPORTING VALUES, GROUPS, AND COMPARISONS IN TABLES

*"Tables are for communication,*
*not data storage."*[1]

## Introduction

In this and the next chapter, we present guidelines for displaying information in tables in graphs according to 3 areas of emphasis: 1) values, 2) groups of related values, and 3) comparisons between 2 or more groups. In designing tables, these 3 areas consist of the following:

- *Values*: the information in an individual cell in a table; a single datum, symbol, or observation.

- *Groups*: the columns and rows in a table, which contain values either from the same group or of the same class. Included here are descriptive statistics summarizing groups or distributions of data, such as a total, percentage, mean and standard deviation, or median and interquartile range.

- *Comparisons*: columns and rows that summarize or compare 2 or more groups. In this level are columns and rows showing differences between groups, totals for 2 or more groups; correlation coefficients; odds, risk, and hazards ratios; estimates and confidence intervals; and P values.

## Functions of Tables

Tables function well to:

- Organize and data, especially exact numbers, more clearly and concisely than can be done with words [2]

- Condense or summarize large amounts of data, especially complex or detailed data[2-5]

• Compare individual values or groups of data[2-5]

• Improve the ease and speed with which specific information can be located and understood[4,5]

• Facilitate calculations[4]

## Components and Types of Tables

Tables in scientific publications usually have at least the first 6 components listed below, and many have 7 or all 8 (**Table 17.1**):

1. **Table number** (Exceptions are tables "embedded" in the text, such as lists and simple enumerations, and publication styles in which only 2 or more tables are numbered)

2. **Table title**, which identifies the data displayed in the table and the context in which they should be interpreted

3. **Column** (box) **headings** that identify the information contained in each column, and perhaps "spanner," "straddle," or "decked" headings that group 2 or more columns of related information

4. **Row** (stub) **headings** that identify the information contained in each row, and perhaps subheadings and "cut-in" headings that group 2 or more rows of related information

5. **Data** (in the "data field"); the cells of the table other than those containing column and row headings

6. **Horizontal lines** (rules): usually at least 3: below the title, below the column heads, and below the data field; other appropriate lines include those above totals and those separating major divisions of the data field (above "cut-in" headings)

7. **Expanded forms of abbreviations used in the table**, below the data field

8. **Footnotes referenced in the table**, below the expanded abbreviations, usually used in the following order: *, †, ‡, §, ||, ¶, **, ††, and so on[6] (sometimes the # mark is used before doubling the marks[2]) and some style manuals use superscript lowercase letters[3].

**Table 17.1 [Table number] Components and nomenclature of tables [table title]**

**[rule]**

| **Row Heading** | **Spanner Heading** | | | | **Column Heading,** |
|---|---|---|---|---|---|
| | **[rule]** | | | | |
| **[rule]** | **Column heading, Group size (units)** | **Column heading, Group size (units)** | **Column heading, Group size (units)** | **Column heading, Group size (units)** | **Group size (units)** |
| | | | **"Cut-in" Heading** | | |
| **Row heading** | | | | | |
| | | | **Data Field** | | |
| | **[rule]** | | | | |
| **Row subhead** | | | | | |
| **Row subhead** | | | | | |
| **Row heading** | | | | | |
| | | | **"Cut-in" Heading** | | |
| **Row heading** | | | | | |
| | | | **Data Field** | | |
| | **[rule]** | | | | |
| **Row subhead** | | | | | |
| **Row subhead** | | | | | |
| **Total** | | | | | |
| **[rule]** | | | | | |

**ABC = expanded abbreviation**
***** **(asterisk)**
† (dagger)
**‡** **(double dagger)**
**§** **(section mark)**
**||** **(parallel mark)**
**¶** **(paragraph mark)**
****** **(double asterisk)**

[a] **superscript lowercase letter**
[b] **superscript lowercase letter**
[c] **superscript lowercase letter**

Tables are often referred to by the number of explanatory variables they present. Thus, a 1-way table has only a single explanatory variable, a 2-way table has 2, and a multi-way has 3 or more (**Tables 17.2** to **17.4**). Tables may also be referred to by the number of cells in the data field.  For example, a 2 x 3 table indicates a data field of 6 cells (**Tables 17.2** to **17.4**).

**Table 17.2 A 1-way table shows 1 explanatory variable (here, group, with values of treatment and control).** Also called a 2 x 3 table because the data field contains 6 cells

| Response Variable | Control Group (n = 118) | Treatment Group (n = 123) |
|---|---|---|
| Variable 1, mg | | |
| Variable 2, kg | | |
| Variable 3, mg/dL | | |

**Table 17.3 A 2-way table shows 2 explanatory variables (group, with values of treatment and control, and sex of patients in each group).** Also called a 4 x 3 table because the data field contains 12 cells. The column headings of control and treatment group are now "spanner headings that cover the subgroups of men and women.

| Response Variable | Control Group (n = 118) | | Treatment Group (n = 123) | |
|---|---|---|---|---|
| | Men (n = 57) | Women (n = 61) | Men (n = 55) | Women (n = 63) |
| Variable 1, mg | | | | |
| Variable 2, kg | | | | |
| Variable 3, mg/dL | | | | |

**Table 17.4 A 3-way table shows 3 explanatory variables (group, with values of treatment and control; sex of patients in each group; and handedness, with values of right and left).** Also called a 8 x 3 table because the data field contains 24 cells. The column headings of men and women are now themselves spanner headings that cover the subgroups of right- and left-handedness.

| Response Variable | Control Group (n = 118) | | | | Treatment Group (n = 123) | | | |
|---|---|---|---|---|---|---|---|---|
| | Men (n = 57) | | Women (n = 61) | | Men (n = 57) | | Women (n = 66) | |
| | Right (n = 45) | Left (n = 12) | Right (n = 48) | Left (n = 13) | Right (n = 47) | Left (n = 10) | Right (n = 51) | Left (n = 15) |
| Variable 1, mg | | | | | | | | |
| Variable 2, kg | | | | | | | | |
| Variable 3, mg/dL | | | | | | | | |

## Principles of Table Construction

If tables are to communicate information quickly and accurately, readers must be able to 1) identify how the information is organized, 2) find the information of interest, and 3) interpret the information once it is found.[7] Experience, convention, and some research studies have identified at least 6 principles that should guide the construction of tables:

**1. Tables should have a purpose; they should contribute to and be integrated with the rest of the text.** [1,2, 8, 9]

Data should not be reported for their own sake. Rather, they should be part of a larger effort to answer the four questions of research: "What did you do?," "Why did you do it?," "What did you find?," and "What does it mean?". Thus, tables should be used only when they can communicate information more efficiently or effectively than can be done in text or figures.

**2. The purpose of the table should determine its form.**[4,5,10]

A table created to collect data is not necessarily the same table that should be used to communicate these data. A table created to organize a large amount of data so that a value can be referenced easily will not necessarily be the same as a table constructed to emphasize patterns in the data or comparisons among the patterns.

Tables may be structured for analytical or reference functions.[5] Analytical tables are designed "from the inside out" by organizing the data field to help reveal patterns in the data. Reference tables are designed "from the outside in" by organizing the column and row heads to help readers find specific information quickly. For example, **Tables 17.5** and **17.6** present the same information but **Table 17.5** lists the causes of death from high to low to focus attention on the relative number of deaths from each type of cancer. **Table 17.6** lists the types of cancer alphabetically to help readers find information on a specific type of cancer faster.

**Table 17.5 Leading Causes of Cancer Death in the US, 1998**

| Type of Cancer | Number of Deaths |
|---|---|
| **Lung** | 160,000 |
| **Colorectal** | 57,000 |
| **Breast** | 44,000 |
| **Prostate** | 39,000 |
| **Cervical** | 5,000 |

**Table 17.6 Leading Causes of Cancer Death in the US, 1998**

| Type of Cancer | Number of Deaths |
|---|---|
| **Breast** | 44,000 |
| **Cervical** | 5,000 |
| **Colorectal** | 57,000 |
| **Lung** | 160,000 |

| | |
|---|---|
| **Prostate** | 39,000 |

**3. Tables should be organized and formatted to assist readers in finding, seeing, understanding, and remembering the information**.

A table that contains all the necessary data but forces readers to organize the data before understanding it does everyone a disservice: it increases the time needed for readers to evaluate the data and does not assure that the author's understanding of the data will match the readers' interpretation of it.

The data in **Table 17.7** are probably organized in the order in which they were collected. Ordering the columns by patient number conveys no information to readers. In addition, the characteristics studied are not organized in any particular way. In contrast, in **Table 17.8**, the characteristics are grouped by category to add structure, and the columns are ordered first by sex and then by age. New patient numbers easily identify the first 4 patients as men and the next 4 as women.

**Table 17.7 Baseline Fitness Data for the 8 Patients Completing the Fitness Training Study, as Collected.**

| **Characteristic** | **Patient Number** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **Age, y** | 35 | 16 | 21 | 19 | 41 | 30 | 22 | 37 |
| **Sex** | M | M | F | F | M | F | M | F |
| **Resting pulse, beats/min** | xx | xx | xx | xx | xx | xx | xx | xx |
| **1.5-mile run, min** | xx | xx | xx | xx | xx | xx | xx | xx |
| **Hemoglobin, g/dL** | xx | xx | xx | xx | xx | xx | xx | xx |
| **Weight, kg** | xx | xx | xx | xx | xx | xx | xx | xx |
| **$10^6$ x RBCs /μL** | xx | xx | xx | xx | xx | xx | xx | xx |

RBC = red blood cells

**Table 17.8 Baseline Fitness Data for the 8 Patients Completing the Fitness Training Study, Revised.**

| | **New Patient Number [Old patient number, for illustration only]** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Men** | | | | **Women** | | | |
| | **1 [3]** | **2 [1]** | **3 [6]** | **4 [5]** | **5 [4]** | **6 [7]** | **7 [2]** | **8 [8]** |
| **Age, y** | 16 | 22 | 35 | 41 | 19 | 21 | 30 | 37 |
| **Weight, kg** | xx | xx | xx | xx | xx | xx | xx | xx |
| **Blood values** | | | | | | | | |
| Hemoglobin, g/dL | xx | xx | xx | xx | xx | xx | xx | xx |
| $10^6$ x RBCs /µL | xx | xx | xx | xx | xx | xx | xx | xx |
| **Fitness Results** | | | | | | | | |
| 1.5-mile run, min | xx | xx | xx | xx | xx | xx | xx | xx |
| Resting pulse, beats/min | xx | xx | xx | xx | xx | xx | xx | xx |

RBC = red blood cells

## 4. Values to be compared should be placed side-by-side.[1, 2]

English is read from left to right and from top to bottom. Thus, at least in English language publications, placing values side-by-side is not only the easiest way to compare them, but it also encourages this comparison. In biomedical research, where a treatment group is compared to a control group, values for each group should be given in adjacent columns so that the variables in each row can be compared more easily.

**Table 17.9** reports the effect of physician experience on surgical simulation scores. That is, the author wants to show the relationship between the explanatory variable of experience and the response variable of surgical skill, as measured here by performance on simulation tasks. The table is poorly organized to show this relationship, however. Placing the simulation scores side-by-side encourages readers to compare individual surgical skills within each level of experience (**Table 7.10)**

**Table 17.9 Effect of Physician Experience on Surgical Simulation Scores**

| **Experience** | **Surgical Simulation Scores*** | | |
|---|---|---|---|
| | **Cutting** | **Suturing** | **Ablating** |
| **Residents (n = 12)** | 79 | 63 | 80 |
| **Fellows (n = 8)** | 88 | 87 | 91 |
| **Staff Surgeons (n = 15)** | 96 | 92 | 97 |
| **P†** | 0.03 | 0.004 | 0.05 |

* Scores range from a low of 0 to a high of 100
† ANOVA. Differences were significant only between residents and staff surgeons

**Table 17.10 Effect of Physician Experience on Surgical Simulation Scores**

| **Surgical Simulation** | **Simulation Scores (low = 0; high = 100)** | | | **P†** |
|---|---|---|---|---|
| | **Residents (n = 12)** | **Fellows (n = 8)** | **Staff Surgeons (n = 15)** | |
| **Cutting** | 79 | 88 | 96 | 0.03 |
| **Suturing** | 63 | 87 | 92 | 0.004 |
| **Ablating** | 80 | 91 | 97 | 0.05 |

† ANOVA. Differences were significant only between residents and staff surgeons

**5. Organize the table visually as well as functionally.**[3]

Graphic elements, including spacing, should be used to help organize the table visually. Elements such as lines, bold type, outlined cells, spacing, and shading, can help readers make within- and between-group comparisons, differentiate more-important values from less-important ones, highlight patterns in the data, indicate special circumstances associated with the data, and so on. For example, Table **17.11** supports the argument that a single exposure to ultraviolet light killed all the cell lines in a 24-hour study. A space inserted after every 5th cell line makes the list easier to read, shaded columns indicate the treatment period, and the relationship between the treatment and outcome is supported visually.

Most publishers specify their format for tables, and some may not allow the full use of design elements as recommended here.

**Table 17.11 Number of cells before and after ultraviolet treatment over 24 hours.** Shaded columns indicate the hours under ultraviolet light, and outlined cells indicate the time at which the cell counts reached zero. (The last 3 columns could be deleted because the values do not vary. Here, they simply indicate the data collection periods and the data recorded in them.)

| **Cell line** | **Time of Data Collection (Cell Counts x $10^3$/mL)** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **9 am** | **12 pm** | **3 pm** | **6 pm** | **9 pm** | **12 am** | **3 am** | **6 am** | **9 am** |
| **1** | 5.4 | 5.6 | 5.6 | 0 | 0 | 0 | 0 | 0 | 0 |
| **2** | 3.2 | 3.2 | 3.2 | 1.1 | 0 | 0 | 0 | 0 | 0 |
| **3** | 12.3 | 13.1 | 13.3 | 0 | | 0 | 0 | 0 | 0 |
| **4** | 6.0 | 5.5 | 5.1 | 1.2 | 0.04 | 0 | 0 | 0 | 0 |
| **5** | 24.3 | 25.0 | 26.2 | 12.9 | 0 | 0 | 0 | 0 | 0 |
| | | | | | | | | | |
| **6** | 17.8 | 18.1 | 18.7 | 0 | 0 | 0 | 0 | 0 | 0 |
| **7** | 14.2 | 14.0 | 14.5 | 0 | 0 | 0 | 0 | 0 | 0 |
| **8** | 76.0 | 75.0 | 76.0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **9** | 78.9 | 82.7 | 83.2 | 42.9 | 12.3 | 0 | 0 | 0 | 0 |
| **10** | 34.5 | 31.2 | 33.0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | | | | | | | | |
| **11** | 23.7 | 26.3 | 24.0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **12** | 49.2 | 50.8 | 49.6 | 0 | 0 | 0 | 0 | 0 | 0 |

**6. Data presented in tables should not be duplicated in the text.**[3]

Describing in the text data that are also presented in a table remains a common problem, even though most style guides and journals advise against the practice. Duplicate information takes valuable space and so is to be avoided in print publications.

Values, groups, or comparisons in tables can be mentioned in the text, of course, but the table should present the data.

Tables should also be kept as simple as possible.[3] Include only the information that is relevant to the purpose of the table.

## Guidelines for Writing Table Titles

### Guideline 17.1 Whenever possible, the title and the table should allow the data to be understood without reference to the text.[3,9]

At least in scientific publications, tables and figures are often separated from their associated text. In such cases, the data can be uninterpretable because the context in which they need to be understood is no longer available. For this reason, the title and structure of the table should supply enough context so that the table can stand alone. However, the title should not provide detailed background information or summarize or interpret the results[2]; these explanations are best given in the text.

When several related tables are presented together, the context of the study should be included in the first, but it need not be repeated in the others if such repetition becomes tedious or requires too much space.

Poor Title: Too general

**Table 12. Patient Characteristics**

Better Title: More specific

**Table 12. Baseline Characteristics of 32 Patients with Malignant Hematological Disease Treated with G-CSF-Primed Bone Marrow Cells**

Poor Title: Incomplete

**Table 5. Akahori's Classification for Staging Preoperative Status**

Better Title: Complete

**Table 5. Akahori's Classification for Staging Preoperative Status of Hand Surgery Patients**

### Guideline 17.2 The title should identify the data shown in the data field.[3]

When writing a table title, begin by identifying the data in the table. Avoid simply repeating the column and row heads.

Poor Title: Needlessly repeats the column and row heads

**Table 2. Mean Test Scores for Knowledge, Recall, and Satisfaction for Intervention and Control Groups after Exposure to Printed, Multi-Media, Interpersonal, or Audio Self-Care Instructions**

Better Title: Better describes the data presented

**Table 2. Self-Care Effectiveness Scores for 1,472 Patients Receiving Self-Care Instructions through Different Modes of Communication**

## Guidelines for Writing Column and Row Heads

### Guideline 17.3 Column and row headings should use terms that appear in the text.[3,10]

Readers enter a table through column and row headings; thus, familiar terms in headings (terms also used in the text) are usually more effective than unfamiliar terms.[10] The information, symbols, and units of measurement in a table should also be consistent with those in the text.[3] Every column and every row should have a heading.[3]

▲ A common source of confusion is the inconsistent use of terms for study groups anywhere in the text. For example, authors may speak of a study's participants, patients, survivors, individuals, subjects, volunteers, and so on. They may speak of the treatment, intervention, or active group and of the placebo, control, untreated, or inactive group. Consistency is usually more important than the term chosen.

### Guideline 17.4 Use "spanner" and "cut-in" headings across columns and row headings above rows to identify subgroups.

Spanner and row heads clarify the organization of data (**Tables 17.12** and **17.13**). Each level of heading adds an additional variable to the table, however, which can quickly make the table more crowded and complex as columns are added.

In some cases, row headings can be expanded into "cut-in" headings that span all the columns. Cut-in headings emphasize subgroups and essentially create two or more tables with identical column heads (**Table 17.1**).

**Table 17.12 A table in need of a spanner head.**

| Variable | Low Dose | High Dose | Low Dose | High Dose |
|---|---|---|---|---|
| Variable #1 | | | | |
| Variable #2 | | | | |
| Variable #3 | | | | |

**Table 17.13 A table showing the value of an added spanner head.**

| Variable | Control Group | | Treatment Group | |
|---|---|---|---|---|
| | Low Dose | High Dose | Low Dose | High Dose |
| Variable #1 | | | | |
| Variable #2 | | | | |
| Variable #3 | | | | |

To conserve space, information on 2 or more variables can sometimes be combined in a single cell.[5] For example, rather than 1 column for age and 1 for sex, both can be indicated in a single column, appropriately labeled, as, say, "f/34," to indicate a 34-year-old woman.

## Guideline 17.5 When applicable, report group size, units of measurement, or both, in the column or row headings.[3,9]

In addition to the name of the variable, the group size and units of measurement are essential for interpreting the data and for checking the consistency of the data. In **Table 17.14**, height could be measured in inches or centimeters, weight in pounds or kilograms, and temperature in degrees Fahrenheit or degrees Celsius. In addition, without the number of patients in each of the dosage groups, the number of observations represented by the data has to be derived from the text. In **Table 17.15**, these problems have been eliminated.

**Table 17.14**

| Variable | Control Group | | Treatment Group | |
|---|---|---|---|---|
| | Low Dose | High Dose | Low Dose | High Dose |
| **Height** | | | | |
| **Weight** | | | | |
| **Temperature** | | | | |

**Table 17.15**

| Variable | Control Group (n = 29) | | Treatment Group (n = 27) | |
|---|---|---|---|---|
| | Low Dose (n = 13) | High Dose (n = 16) | Low Dose (n = 12) | High Dose (n = 15) |
| **Height, cm** | | | | |
| **Weight, kg** | | | | |
| **Temperature, °C** | | | | |

▲ When reporting measurements with multipliers, use the form "x $10^3$ mg" rather than "mg (x $10^3$)." Placing the multiplier in front of the unit indicates that the values in the cells are already in the multiplied form, whereas placing it after the unit may be interpreted to mean that the reader must do the multiplication.[3] Thus, a cell showing "15 mg (x $10^3$)" could be interpreted as reporting a value of "15,000 mg," if the reader is expected to expand the notation, or of "0.015 mg," if the notation has already been expanded by the author, whereas "15 x $10^3$ mg" always means "15,000 mg."

## Guidelines for Presenting Individual Values

### Guideline 17.6 Round numbers to 2 significant digits, unless additional precision is necessary.[1]

As discussed in Guideline 2.1, most readers can deal effectively with numbers of only 2 significant digits. This circumstance holds for numbers in tables as well (**Tables 17.16**).

Sometimes, it is desirable to report data to the level of precision of the measurement.[2] Means and other calculated values can be reported to one additional significant digit.[2] Thus, when a table is designed to present actual values for reference, readers can see not only the values but also the precision of the measurements. However, when the primary purpose of a table or a slide is to communicate patterns or comparisons, rounding to 2 significant digits is preferred.

**Table 17.16 In addition to helping readers understand information, rounding reduces the visual complexity of a table. Numbers should be rounded unless there is a compelling reason to report them with more precision; however, numbers should be rounded only for communication, not for analysis.**

| Variable | Control Group (n = 20) | | Treatment Group (n = 20) | |
|---|---|---|---|---|
| | Low Dose (n = 10) | High Dose (n = 10) | Low Dose (n = 10) | High Dose (n = 10) |
| **Unrounded Numbers are often Unnecessarily Precise** | | | | |
| **Age, y** | 35.97 | 16.34 | 21.12 | 19.04 |
| **Weight, kg** | 61.43 | 81.57 | 58.83 | 100.67 |
| **Rounded Numbers are More Quickly Read and More Easily Remember ed** | | | | |
| **Age, y** | 36 | 16 | 21 | 19 |
| **Weight, kg** | 61 | 82 | 59 | 100 |

## Guideline 17.7 Align data, symbols, and text consistently.[3]

Tables are more easily read when the data or symbols in the cells are presented in a consistent visual form. Once the form is learned, it is applied to the rest of the table, making interpretation faster and easier. Proper alignment can also give readers an additional visual cue about the magnitude of a number.

Elements on which numbers are usually aligned are decimal points, plus or minus signs, hyphens (used to indicate ranges), parentheses, virgules (slashes: /) and factors of 10 (**Tables 17.17** and **17.18**). Symbols and text in cells may be centered or aligned on the left margin ("flush left" margins). Justified margins are not recommended for words, in either tables or text, because the justification creates uneven spacing between words that can interfere with reading. "Flush-right" margins are also not recommended.

**Table 17.17 Examples of Poorly Aligned Data, Symbols, and Text in a Table**

| Poor Alignment on Decimal Points | Poor Alignment on Parentheses Mean (SD) | Poor Alignment on factors of 10 | Uncentered Symbols | Justified Margins |
|---|---|---|---|---|
| 2.81 | 12 (6) | 23 x $10^3$ | ↓ (too high) | This text is not set with a flush left margin |
| 143.5 | 762 (51) | 5567 x $10^3$ | (too low) ↑ | This text is set with justified margins |
| 3.687 | 39,453 (321) | 9.8 x $10^4$ | (too low) ↔ | This text is not centered between margins |

**Table 17.18 Examples of Well Aligned Data, Symbols, and Text in a Table**

| Alignment on Decimal Points | Alignment on Parentheses Mean (SD) | Alignment on factors of 10 | Centered Symbols | Margins |
|---|---|---|---|---|
| 2.81 | 12 (6) | 23 x $10^3$ | ↓ | This text is set with a flush left margin |
| 143.5 | 762 (51) | 5567 x $10^3$ | ↑ | This text is set with a flush-right margin (not recommended) |
| 3.68 | 3453 (321) | 9.8 x $10^4$ | ↔ | This text is centered between margins |

## Guideline 17.8 Do not leave empty cells in a table unless an entry would be illogical.[3]

Empty cells create ambiguity because there is no assurance that data have been accidentally omitted from them. For this reason, some indication that the cell does not contain data is essential. One solution is to fill the cell with an abbreviation, such as NA for “not available” or “not applicable” or nd for “not detectable,” “not done,” “or not determined,” and to expand these abbreviation below the data field. Other messages can also be used, such as “data not calculated,” “data lost,” and so on. Another solution is to fill the empty cell with ellipses (...) or the longer "em-dash" (—) and to footnote the reason for the empty cell.[2,3]

Tables that may contain empty cells are those reporting paired data or correlation matrices (**Table 6.2**). These tables have duplicate cells because the column and row heads are the same. A good example is a mileage chart. The names of cities are the column headings as well as the row headings. One can follow the row for Los Angeles to the column for New York and find the distance between them. However, one could also follow the row for New York to the column for Los Angeles. To simplify the table, only 1 of the 2 combinations will contain data, leaving half the cells empty.

▲ Empty cells should not be interpreted as having a value of zero.

▲ Do not use the abbreviation "NS" for "not statistically significant." Instead, report the actual P value.[2] If the relationship is important enough to test, the result is important enough to report. (*See* **Guideline 4.15**.)

### Guideline 17.9 Highlight important values[8]

If intelligence is "knowing what to attend to," readers can be made "more intelligent" by guiding their attention to more important relationships and away from less important ones. Highlighting individual cells or groups of cells that contain important data or patterns helps readers know what they should be attending to (**Tables 17.11**). Outlining the cell, presenting the contents of the cell in bold type, or shading are common ways to emphasize important values.

## Guidelines for Presenting Groups of Values

### Guideline 17.10 Order the rows and columns sensibly[2,3]

The order in which column and row heads appear organizes the table and, as such, can clarify or obscure patterns in the data or make finding information easier or more difficult (**Tables 17.5** and **17.6**). Data supporting claims of cause-and-effect or reflecting before-and-after studies should be presented from left to right.

### Guideline 17.11 Keep the data in each cell consistent with both its column and row headings.[2]

Mixing data of different types, different levels of measurement, or different units of measurement in a column or row destroys the consistency that allows readers to rapidly make sense of the table. The most common problem is mixing data from different levels of measurement in the same column. For example, in **Table 17.19**, the data in the columns are inconsistent. The columns are restricted by the column heading to numbers and percent but include data on age, presented as means and standard deviations. Solutions to this problem are shown in **Tables 17.20** and **17.21**.

**Table 17.19 The data integrity of the columns has been violated here by the form of the information on age: mean and standard deviation are reported in a column restricted by its heading to reporting numbers and percents.**

| Variable | Control Group (n = 66) n (%) | Treatment Group (n =83) n (%) |
|---|---|---|
| **Women** | 45 (68) | 54 (656) |
| **Mean (SD) age, y** | 36 (7.3) | 35 (7.0) |
| **Symptomatic** | 19 (29) | 26 (31) |

**Table 17.20 Here, data integrity has been restored by moving the restrictive column head to the row head**

| Variable, units | Control Group (n = 66) | Treatment Group (n =83) |
|---|---|---|
| **Women, n (%)** | 45 (68) | 54 (656) |
| **Mean (SD) age, y** | 36 (7.3) | 35 (7.0) |
| **Symptomatic, n (%)** | 19 (29) | 26 (31) |

**Table 17.21 Another solution to the data integrity problem is to footnote the one or two observations whose form differs from that specified by the column heading.**

| Variable | Control Group* (n = 66) n (%) | Treatment Group† (n =83) n (%) |
|---|---|---|
| **Women** | 45 (68) | 54 (656) |
| **Symptomatic** | 19 (29) | 26 (31) |

* Mean age, years (SD) = 36 (7.3)
† Mean age, years (SD) = 35 (7.0)

### Guideline 17.12 When appropriate, include column and row totals, percentages, or both.

Column and row totals provide not only summary information about the data in the column or row, but they also allow readers to check the numbers for accuracy (**Table 17.22** and **17.23**).

**Table 17.22 Disposition of the Surveys from the First Mailing.** Without column and row totals, readers have to calculate the number of evaluable surveys that form the sample for this study.

| **Characteristic** | **Clinic 1 n (%)** | **Clinic 2 n (%)** | **Clinic 3 n (%)** |
|---|---|---|---|
| **Surveys sent** | 758 (100) | 1,259 (100) | 53 (100) |
| **Undeliverable** | 35 (5) | 79 (6) | 3 (6) |
| **Returned** | 704 (93) | 1,138 (90) | 50 (94) |
| **Incomplete*** | 19 (3)* | 42 (4)* | 1 (2)* |

* As a percent of surveys returned

**Table 17.23 Disposition of the Surveys from the First Mailing.**

| **Characteristic** | **Clinic 1 n (%)** | **Clinic 2 n (%)** | **Clinic 3 n (%)** | **Totals N (%)** |
|---|---|---|---|---|
| **Surveys sent** | 758 (100) | 1259 (100) | 53 (100) | 2,070 (100) |
| **Undeliverable** | 35 (5) | 79 (6) | 3 (6) | 117 (6) |
| **Unreturned** | 19 (3) | 42 (3) | 0 (0) | 61 (3) |
| **Returned Surveys** | 704 (93) | 1138 (90) | 50 (94) | 1,892 (91) |
| **Returned incomplete*** | 19 (3)* | 42 (4)* | 1 (2)* | 62 (3)* |
| **Returned evaluable†** | **685 (90) †** | **1141 (91) †** | **49 (98) †** | **1830 (88) †** |

* As a percent of surveys returned
† As a percent of surveys mailed

### Guideline 17.13 If necessary, number columns or rows to help integrate the text and the table.

A table can sometimes require extensive explanations in the text, and extensive explanations in the text sometimes require frequent reference to a table. In such cases, it can be helpful to number the column or row heads to help readers find their place in the table faster (**Table 17.24**. See also the discussion of age adjustment in Chapter 16).

**Table 17.24 Calculation of Standard Mortality Ratios for Widget Makers, by Age Category.** Numbering the columns allows readers to determine how new values were calculated from those earlier in the table. The text can also refer to these numbers in explaining the calculations further.

| Age Group | [1] Death Rate of Comparison Population, per 1,000 | [2] Total Population of Widget Makers | [3] Expected No. Deaths Among Widget Makers at Comparison Rate ([1] x [2])/1000 | [4] Observed No. Deaths Among Widget Makers | [5] Standard Mortality Ratio, Widget Makers/ Comparison Group [4]/[3] x 100 |
|---|---|---|---|---|---|
| **0 to 19 years** | 9.0 | 8,000 | 72 | 140 | 194 |
| **20 to 49 years** | 5.0 | 12,000 | 68 | 100 | 147 |
| **50+ years** | 4.0 | 13,000 | 60 | 58 | 97 |
| **Total** | **5.6** | **33,000** | **200** | **298** | **149** |

### Guideline 17.14 When appropriate, do not include rows or columns containing values that do not vary.

Columns or rows in which all cells contain identical values may be uninformative. A footnote to the table or a line in the text stating that the values of a given variable did not vary may be a more efficient way to present the data.[2]

## Guidelines for Presenting Comparisons of Values or Groups of Values

### Guideline 17.15 Put data to be compared in adjacent columns.[2]

A table presenting data on 3 variables, say, the sex, age category, and nationality of athletes, can take any of 8 forms, depending on how the column and row heads are arranged (**Table 17.25**). Although each of these 8 forms contains the same information, the table that puts the values to be compared side-by-side is preferred. Thus, Form 1 of **Table 17.25** should be better for comparing the number of male and female athletes from the two countries than, say Form 6. Form 5, on the other hand, would be preferred for comparing the number of athletes within each age group from each country.

Tables are more often limited by width than by length.[3] Thus, sometimes columns and rows must be reversed simply to fit the table on a printed page.

**Table 17.25 A table reporting 3 variables (nationality, sex, and age group) may take any of 8 forms:**

**Form 1**

| | Men | | Women | |
|---|---|---|---|---|
| | US | China | US | China |
| 0 to 21y | | | | |
| 22 to 49y | | | | |
| 50$^{+}$y | | | | |

**Form 2**

| | China | | US | |
|---|---|---|---|---|
| | Men | Women | Men | Women |
| 0 to 21y | | | | |
| 22 to 49y | | | | |
| 50$^{+}$y | | | | |

**Form 3**

| | 0 to 21y | | 22 to 49y | | 50$^{+}$y | |
|---|---|---|---|---|---|---|
| | Men | Women | Men | Women | Men | Women |
| US | | | | | | |
| China | | | | | | |

**Form 4**

| | Men | | | Women | | |
|---|---|---|---|---|---|---|
| | 0 to 21y | 22 to 49y | 50$^{+}$y | 0 to 21y | 22 to 49y | 50$^{+}$y |
| US | | | | | | |
| China | | | | | | |

**Form 5**

| | 0 to 21y | | 22 to 49y | | 50$^{+}$y | |
|---|---|---|---|---|---|---|
| | US | CHINA | US | China | US | China |
| Men | | | | | | |
| Women | | | | | | |

**Form 6**

| | US | | | China | | |
|---|---|---|---|---|---|---|
| | 0 to 21y | 22 to 49y | 50$^{+}$y | 0 to 21y | 22 to 49y | 50$^{+}$y |
| Men | | | | | | |
| Women | | | | | | |

**Form 7**

| | Age | | |
|---|---|---|---|
| | 0 to 1y | 22 to 49y | 50$^{+}$y |
| **MEN** | | | |
| US | | | |
| China | | | |
| **WOMEN** | | | |
| US | | | |
| China | | | |

**Form 8**

| | Age | | |
|---|---|---|---|
| | 0 to 1y | 22 to 49y | 50$^{+}$y |
| **US** | | | |
| Men | | | |
| Women | | | |
| **CHINA** | | | |
| Men | | | |
| Women | | | |

## Guideline 17.16 When appropriate, include columns or rows with summary or inferential statistics, especially estimates and confidence intervals.

Two or more groups are often compared with summary statistics. For example, columns or rows can be included to present differences between groups, estimates and confidence intervals, and P values. Many authors compare two groups by giving only their mean values and the P value for the difference between the means (**Table 17.26**). The message of such a table is that "the groups are statistically significantly different." A more effective way to present this comparison is to include a column showing the difference between means (the estimate) and the confidence interval for this estimate (**Table 17.27**). Readers can then judge the clinical importance of the difference (the size of the treatment effect), and can determine whether the confidence interval includes clinically unimportant values that would indicate an inconclusive study. (See Chapter 3 on Confidence Intervals.)

**Table 17.26 Test Results for 345 Patients with Esophageal Varices Treated with Laser Coagulation.** The table emphasizes the P value at the expense of the estimated treatment effect and its 95% confidence interval.

| Variable | Treatment Mean (SD) | Control Mean (SD) | P* |
|---|---|---|---|
| **Test #1 score** | 67 (21.5) | 52 (19.8) | 0.01 |
| **Test #2 score** | 24 (3.0) | 27 (2.3) | 0.08 |
| **Test #3 score** | 89 (9.1) | 48 (8.6) | 0.002 |

* Student's *t* test

**Table 17.27 Test Results for 345 Patients with Esophageal Varices treated with Laser Coagulation.** The table emphasizes the estimated treatment effect and its 95% confidence interval, which are more clinically meaningful than P values.

| Variable | Treatment Mean (SD) | Control Mean (SD) | Difference (95% CI) | P* |
|---|---|---|---|---|
| **Test #1 score** | 67 (21.5) | 52 (19.8) | 15 (3.5 to 26.5) | 0.01 |
| **Test #2 score** | 24 (3.0) | 27 (2.3) | 3 ( to 5.0 to 11.2) | 0.08 |
| **Test #3 score** | 89 (9.1) | 48 (8.6) | 41 (35.6 to 46.4) | 0.002 |

* Student's *t* test

## Guideline 17.17 Highlight unusual or important values to be compared across groups.

Calling attention to group values of interest, such as maximum, minimum, or atypical values, helps readers focus their attention appropriately (**Table 17.11**).

**Guideline 17.18 When presenting the same information about different groups in different tables, use tables with identical formats.[3]**

Space limits often require dividing a table into 2 or more separate tables. For example, when treatment and control groups (named in column headings) are being compared on several outcomes (listed in row headings), the table may become too long, depending on the number of outcomes listed. In such cases, presenting the outcomes as a series of tables, one, say, for neurological outcomes, one for functional outcomes, and one for quality to of to life outcomes, may be useful. In such cases, keeping the formats of the individual tables identical makes the data more available to readers.

## Standard Tables Commonly Used in Reporting Biomedical Research

Many tables used to report biomedical data are standard and need not be reinvented each time they occur. Below are listed many of the more common of these tables that appear in this book, with their associated table number and page number.

| Table | Table Number | Page Number |
|---|---|---|
| Contingency tables for chi-square analyses | 6.1 | |
| Correlation matrices | 6.2 | |
| Reporting multiple regression models | 7.1 | |
| Reporting simple logistic regression models | 7.2 | |
| Reporting multiple logistic regression models | 7.3 | |
| Reporting ANOVA models | 8.1, 8.2 | |
| Reporting Kaplan-Meier estimates | 9.1 | |
| Life tables | 9.2 | |
| Reporting Cox proportional hazards models | 9.3 | |
| Tables for computing diagnostic test characteristics | 10.1 | |

## Acknowledgments

We thank Jessica Ancker, MPH, Adam Jacobs, PhD, Cassandra Talerico, Almost a PhD, Barbara Gastel, MD, and Dan Liberthson, PhD, for their careful review and thoughtful comments on this chapter.

## References

1. Wainer H. Understanding graphs and tables. Ed Researcher 1992;21:14-23.

2. American Medical Association. American Medical Association Manual of Style: A Guide for Authors and Editors, 9th ed. Baltimore: Williams & Wilkins, 1998.

3. Style Manual Committee, Council of Biology Editors. Scientific Style and Format: The CBE Manual for Authors, Editors, and Publishers 6th ed. Council of Biology Editors (now the Council of Science Editors). Cambridge: Cambridge University Press, 1994.

4. Briscoe MH. Preparing Scientific Illustrations: A Guide to Better Posters, Presentations, and Publications. 2nd edition. New York: Springer-Verlag, 1996.

5. Harris RL. Information Graphics: A Comprehensive Illustrated Reference. Oxford: Oxford University Press, 1999.

6. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals. 2001 update. http://www.icmje.org/

7. Wright P. A user-oriented approach to the design of tables and flowcharts. In: Jonassen DH, editor. The Technology of Text: Principles for Structuring, Designing, and Displaying Text: vol 1. Englewood Cliffs, New Jersey: Educational Technology Publications, 1982:317-40. Cited in: Schriver KA. Dynamics in Document Design. New York: Wiley Computer Publishing, 1997.

8. White J. Using Charts and Graphs: 1000 Ideas for Visual Persuasion. New York: R. R. Bowker Company, 1984.

9. Jordan EP, Shepard WC. Rx for Medical Writing. Philadelphia: W. B. Saunders Company, 1952.

10. Wright P.  Presenting technical information: a survey of research findings. Instruct Sci 1977;6:93-134

## Guidelines for Writing Table Titles

| | |
|---|---|
| Guideline 17.1 | Whenever possible, the title and the table should allow the data to be understood without reference to the text. |
| Guideline 17.2 | The title should identify the data shown in the data field. |

## Guidelines for Writing Column and Row Heads

| | |
|---|---|
| Guideline 17.3 | Column and row headings should use terms that appear in the text. |
| Guideline 17.4 | Use "spanner" and "cut-in" headings across columns and row headings above rows to identify subgroups. |
| Guideline 17.5 | When applicable, report group size, units of measurement, or both, in the column or row headings. |

## Guidelines for Presenting Individual Values

| | |
|---|---|
| Guideline 17.6 | Round numbers to 2 significant digits, unless additional precision is necessary. |
| Guideline 17.7 | Align data, symbols, and text consistently. |
| Guideline 17.8 | Do not leave empty cells in a table unless an entry would be illogical. |
| Guideline 17.9 | Highlight important values |

## Guidelines for Presenting Groups of Values

| | |
|---|---|
| Guideline 17.10 | Order the rows and columns sensibly[2,3] |
| Guideline 17.11 | Keep the data in each cell consistent with both its column and row headings. |
| Guideline 17.12 | When appropriate, include column and row totals, percentages, or both. |
| Guideline 17.13 | If necessary, number columns or rows to help integrate the text and the table. |
| Guideline 17.14 | When appropriate, do not include rows or columns containing values that do not vary. |

## Guidelines for Presenting Comparisons of Values or Groups of Values

| | |
|---|---|
| Guideline 17.15 | Put data to be compared in adjacent columns.[2] |
| Guideline 17.16 | When appropriate, include columns or rows with summary or inferential statistics, especially estimates and confidence intervals. |
| Guideline 17.17 | Highlight unusual or important values to be compared across groups. |
| Guideline 17.18 | When presenting the same information about different groups in different tables, use tables with identical formats.[3] |